

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**Best Available**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

J1046 U.S. PRO  
09/879932  
06/14/01

In re Application of: :  
Maki Yamada et al. :  
Serial No.: To be assigned : Art Unit: To be assigned  
Filed: Herewith : Examiner: To be assigned  
For: METHOD AND APPARATUS : Atty Docket: 0402/00624  
FOR PRODUCTING  
ACOUSTIC MODEL :  
:

**SUBMISSION OF CERTIFIED PRIORITY DOCUMENT(S) and**  
**CLAIM TO PRIORITY UNDER 35 U.S.C. § 119**

Commissioner for Patents  
Washington, D.C. 20231

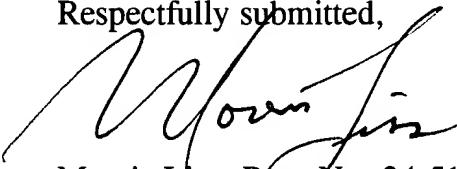
Sir:

Priority under 35 U.S.C. § 119 is hereby claimed to the following priority document(s), certified copies of which are enclosed. The documents were filed in a foreign country within the proper statutory period prior to the filing of the above-referenced United States patent application.

<u>Priority Document Serial No.</u>	<u>Country</u>	<u>Filing Date</u>
2000-194196	Japan	June 28, 2000
_____	_____	_____
_____	_____	_____

Acknowledgement of this claim and submission in the next official communication is respectfully requested.

Respectfully submitted,



Morris Liss, Reg. No. 24,510  
Connolly Bove Lodge & Hutz LLP  
1990 M Street, N.W.  
Washington, D.C. 20036-3425  
Telephone: 202-331-7111

Date: 6/14/01

日本国特許庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

11046 U.S. PTO  
09/879932  
06/14/01



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application: 2000年 6月28日

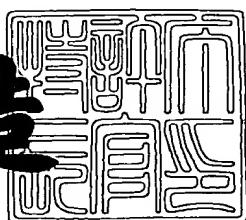
出願番号  
Application Number: 特願2000-194196

出願人  
Applicant(s): 松下電器産業株式会社

2001年 3月16日

特許庁長官  
Commissioner,  
Patent Office

及川耕造



出証番号 出証特2001-3019153

【書類名】 特許願

【整理番号】 2931010223

【提出日】 平成12年 6月28日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/06

【発明者】

【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技研株式会社内

【氏名】 山田 麻紀

【発明者】

【住所又は居所】 神奈川県川崎市多摩区東三田3丁目10番1号 松下技研株式会社内

【氏名】 星見 昌克

【特許出願人】

【識別番号】 000005821

【氏名又は名称】 松下電器産業株式会社

【代理人】

【識別番号】 100097445

【弁理士】

【氏名又は名称】 岩橋 文雄

【選任した代理人】

【識別番号】 100103355

【弁理士】

【氏名又は名称】 坂口 智康

【選任した代理人】

【識別番号】 100109667

【弁理士】

【氏名又は名称】 内藤 浩樹

【手数料の表示】

【予納台帳番号】 011305

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809938

【書類名】 明細書

【発明の名称】 音響モデル作成装置及びその方法

【特許請求の範囲】

【請求項1】 音声認識用の音響モデルを作成する装置において、多数の雑音サンプルをクラスタリングする手段と、各クラスから1つずつ雑音サンプルを選択し学習用複数雑音サンプルとする手段と、その学習用複数雑音サンプルを用いて音響モデルの学習を行う手段とを備えた音響モデル作成装置。

【請求項2】 音声認識用の音響モデルを作成する方法において、多数の雑音サンプルをクラスタリングするステップと、各クラスから1つずつ雑音サンプルを選択し学習用複数雑音サンプルとするステップと、その学習用複数雑音サンプルを用いて音響モデルの学習を行うステップとを備えた音響モデル作成方法。

【請求項3】 プログラムされたコンピュータによって音声認識のための音響モデル作成プログラムを記録した記録媒体であって、多数の雑音サンプルをクラスタリングするステップと、各クラスから1つずつ雑音サンプルを選択し学習用複数雑音サンプルとするステップと、その学習用複数雑音サンプルを用いて音響モデルの学習を行うステップを備えたコンピュータに実行させるためのプログラムを記憶したコンピュータ読み取り可能な記憶媒体。

【請求項4】 請求項1記載の音響モデル作成装置で作成された音響モデルを用いて、入力された音声を認識することを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、雑音環境下でも高い認識率が得られる音声認識用の音響モデル作成装置及びその方法に関する。

【0002】

【従来の技術】

雑音環境下での音声認識において、従来、雑音を重畠させた音声を用いて音響モデルの学習を行っていた（日本音響学会講演論文集、昭和63年3月、3-P-8「ノイズ付加データに対する音素認識システムの評価」）。

【0003】

従来の音響モデル作成装置の構成図を図8に示し、以下に説明する。

【0004】

図8において、201はメモリ、202はCPU、203キーボード/ディスプレイ、204はCPUバス、205は学習用音声サンプル、206は学習用雑音サンプル、207は音響モデルである。

【0005】

図9に示すフローチャートに従って従来の音響モデル作成装置の音響モデルの学習方法について説明する。

【0006】

図9において、Sは各処理ステップを表す。まず、学習用音声サンプル205に学習用雑音サンプル206を重畠し(S81)、一定時間長(以下フレームと呼ぶ。ここでは1フレームは10ミリ秒とする)ごとに音響分析を行い(S82)、音響モデルを学習する(S83)。学習用雑音サンプル206は、会場雑音や車内雑音などを数十秒収録した1種類のデータである。

【0007】

この方法では、認識時の雑音環境と同じような雑音を重畠させた音声で音響モデルの学習を行えば、比較的高い認識率が得られる。

【0008】

【発明が解決しようとする課題】

しかしながら、一般には認識時の雑音環境は未知であり、上記の構成では、音響モデルの学習時の雑音環境と、認識時の雑音環境が異なる場合には認識率の劣化を招くという課題があった。

【0009】

しかし、認識時に出くわし得る雑音すべてを、学習用雑音サンプルとして集めることは不可能である。そこで実際には、ある程度認識時に出くわし得る雑音を想定して、多数の雑音サンプルを集めて学習を行う方法が考えられる。しかし、集めた雑音サンプルすべてに対して音響モデルの学習を行うのは、膨大な時間がかかるため非効率的である。しかも、集めた多数の雑音サンプルの特徴が偏って

いた場合、特徴の偏った雑音サンプルを用いて学習しても、未知の雑音まで広くカバーすることはできない。

【0010】

本発明は、上記従来の課題を解決するもので、比較的少ない複数の雑音を用いて音響モデルの学習を行い、未知の雑音環境下においても高い認識性能が得られるような音響モデルを作成することを目的とする。

【0011】

【課題を解決するための手段】

この課題を解決するために、本発明は、認識時に出くわし得る雑音を想定した多数の雑音サンプルをクラスタリングし、各クラスから1つずつ雑音サンプルを選択し学習用複数雑音サンプルとして学習音声に重畠し、音響モデルを作成する。

【0012】

これにより、未知の雑音環境においても高い認識性能が得られる音響モデルを作成することができる。

【0013】

【発明の実施の形態】

本発明の請求項1に記載の発明は、音声認識用の音響モデルを作成する装置において、多数の雑音サンプルをクラスタリングする手段と、各クラスから1つずつ雑音サンプルを選択し学習用複数雑音サンプルとする手段と、その学習用複数雑音サンプルを用いて音響モデルの学習を行う手段とを備えたものであり、多数の雑音サンプルをクラスタリングして、それぞれのクラスから1つずつ学習に用いる雑音を選択して音響モデルを学習することにより、少ない雑音サンプルで学習でき、しかも偏り無くさまざまな雑音を広くカバーすることができるため、未知雑音環境下においても高い認識性能が得られる音響モデルを作成できるという作用を有する。

【0014】

請求項2に記載の発明は、音声認識用の音響モデルを作成する方法において、多数の雑音サンプルをクラスタリングするステップと、各クラスから1つずつ雑

音サンプルを選択し学習用複数雑音サンプルとするステップと、その学習用複数雑音サンプルを用いて音響モデルの学習を行うステップとを備えたものであり、多数の雑音サンプルをクラスタリングして、それぞれのクラスから1つずつ学習に用いる雑音を選択して音響モデルを学習することにより、少ない雑音サンプルで学習でき、未知雑音環境下においても高い認識性能が得られる音響モデルを作成できるという作用を有する。

【0015】

請求項3に記載の発明は、プログラムされたコンピュータによって音声認識のための音響モデル作成プログラムを記録した記録媒体であって、多数の雑音サンプルをクラスタリングするステップと、各クラスから1つずつ雑音サンプルを選択し学習用複数雑音サンプルとするステップと、その学習用複数雑音サンプルを用いて音響モデルの学習を行うステップを備えた音響モデル作成プログラムをコンピュータに実行することにより、多数の雑音サンプルをクラスタリングして、それぞれのクラスから1つずつ学習に用いる雑音を選択して音響モデルを学習することにより、少ない雑音サンプルで学習でき、未知雑音環境下においても高い認識性能が得られる音響モデルを作成できるという作用を有する。

【0016】

請求項4に記載の発明は、請求項1記載の音響モデル作成装置で作成された音響モデルを用いて、入力された音声を認識するもので、雑音環境下でも高い認識率が得られるという作用を有する。

【0017】

以下、本発明の実施の形態について、図を用いて説明する。

【0018】

(実施の形態1)

図1に、本発明の実施の形態1における音響モデル作成装置の構成図を示し、説明する。図1において、101はメモリ、102はCPU、103はキーボード/ディスプレイ、104はCPUバス、105は学習用音声サンプル、106は多数雑音サンプル、107は学習用複数雑音サンプル、108は音響モデルである。

## 【0019】

本実施の形態では、図1の構成のもとに、最初に図2に示したフローチャートに従って学習用複数雑音サンプルを選択し、次に図3に示したフローチャートに従って音響モデルを作成する。

## 【0020】

最初に、学習用複数雑音サンプルの選択方法について説明する。

## 【0021】

図2において、まずできる限り多くの多数雑音サンプル106を用意する。ここではM個の雑音サンプルがあるものとする。これらM個の雑音サンプルのそれぞれに対し、一定時間長（以下フレームという）ごとに音響分析し（S21）、フレームごとにk次元の特徴パラメータを求める。本実施の形態では、1フレームを10ミリ秒とし、特徴パラメータとして7次までのLPCケプストラム係数（C1, C2, …, C7）を用いる。これを特徴ベクトルと呼ぶ。各雑音ごとにk次元特徴ベクトルの時間平均ベクトルを求める（S22）。

## 【0022】

そして、M個の時間平均特徴ベクトルをクラスタリング手法により、N個のカテゴリーに分類する（S23）。本実施の形態では、クラスタリング手法として、階層的クラスター分析を行う。2つの雑音間の距離は、2つの時間平均特徴ベクトル間の重み付きユークリッド距離を距離尺度とする。2つの雑音間の距離としては他に、ユークリッド距離や、マハラノビスの汎距離、個体間積和、分散を考慮したバッタチャリヤ距離などを用いてもかまわない。2つのクラスター間の距離は、その2つクラスターに属する雑音間の距離の最小値とする（最近隣法）。2つのクラスター間の距離としては他に、2つクラスターに属する雑音間の距離の最大値とする最遠隣法や、2つのクラスターの重心間の距離とする方法や、2つのクラスターに属する個体間の距離の平均とする方法を用いてもかまわない。

## 【0023】

これにより、M個の雑音をN個のクラスに分類することができる。

## 【0024】

本実施の形態ではM=17とし、N=5とする。多数雑音サンプルは図4に示した40秒分の雑音データである。これらを階層的クラスター分析した結果得られる樹形図を図5に示す。

## 【0025】

図5の樹形図では、横方向の長さが個体間あるいはクラスター間の距離をあらわしており、クラスターは樹形図を適当な長さで縦に切断したときに、そこまでで互いに接続している個体によって構成される。図5において☆印で切れば、5つのクラス (N=5) に分類することができる。そこで、

クラス1 {川、音楽}

クラス2 {マークII、カローラ、エスティマ、マジェスタ、ポートピア会場}  
}

クラス3 {データショウ会場、地下鉄}

クラス4 {事業部、営業所、実験室、ざわめき、事務所、町工場}

クラス5 {幼稚園、東京駅}

と定義する。

## 【0026】

そして各クラスから1つずつ任意の雑音を選択し (S24) 、N個の学習用複数雑音サンプル107とする (雑音1~Nとする)。クラス内から1つの雑音を選択する方法としては、クラス内の重心に最も近いものを選択しても良いし、ランダムに選んでも良い。

## 【0027】

ここでは、クラス1から「川」、クラス2から「マークII」、クラス3から「データショウ会場」、クラス4から「事業部」、クラス5から「幼稚園」を選択し、これら雑音を学習用複数雑音サンプル107とする。

## 【0028】

次に、図3に基づいて音響モデルを作成する方法について説明する。

## 【0029】

まず、最初に学習用音声サンプル105に、あらかじめ学習用複数雑音サンプル107に登録されているN個の雑音 (n=1~N) の内1つの雑音を重畠する

(S31)。本実施の形態では学習用音声サンプルとして、音韻バランス単語セット543単語×80名分を音声データとして用いた。雑音の重畠方法は以下に説明する。

## 【0030】

学習用の音声をA/D変換して得られる信号を $S(i)$  ( $i=1, \dots, I$ )、雑音 $n$ をA/D変換して得られる信号を $Nn(i)$  ( $i=1, \dots, I$ )とするとき、この雑音を重畠した雑音重畠音声 $Sn(i)$  ( $i=1, \dots, I$ )は(数1)で表される。ただし、 $I$ はA/D変換時のサンプリング周波数(Hz)にデータの秒数をかけた値である。

## 【0031】

## 【数1】

$$Sn(i) = S(i) + Nn(i) \quad (i=1, \dots, I)$$

## 【0032】

次に、このようにして得られる雑音を重畠した学習用音声サンプルについて、一定時間(フレームと呼ぶ)ごとに音響分析を行う(S32)。ここでは、LPC分析を行いフレームごとにLPCケプストラム係数とその時間回帰係数を求め、特徴パラメータとする。ここではLPCケプストラムを用いるが、FFTケプストラム、MFCC、メルLPCケプストラムなどを用いても良い。このようにして、学習用雑音重畠音声サンプルに対し特徴パラメータの時系列が得られる。

## 【0033】

次に、 $P$ 個の特徴パラメータからなるベクトルを特徴パラメータベクトルとして、音響モデルの学習を行う(S33)。本実施の形態では、特徴パラメータベクトルは1フレーム分の特徴パラメータからなるとするが、複数フレーム分の特徴パラメータを並べて特徴パラメータベクトルとする方法もある。

## 【0034】

学習用複数雑音サンプル107に登録されている $N$ 個の雑音( $n=1 \sim N$ )のすべてについて学習が終了したかを判断し、終了するまでS31～S33の処理を繰り返し、音響モデル108を作成する(S34)。

## 【0035】

音響モデルとしては、DPマッチング用の特徴ベクトルの時系列パターン（標準パターンと呼ぶ）や、HMMなどの確率モデルがある。本実施の形態1では、DPマッチング用の標準パターンを学習するものとする。DPマッチングは時間軸の伸縮を考慮しながら2つのパターンの類似度を計算する効率的な方法である。

#### 【0036】

標準パターンの単位は、一般に音素、音節、半音節、CV/VC(子音+母音、母音+子音)などが用いられる。本実施の形態では音節を標準パターンの単位として説明する。標準パターンのフレーム数は、平均音節フレーム数と同じに設定する。学習用音声サンプルを音節単位に切り出して、DPマッチングにより時間伸縮を考慮しながらフレーム整合を取り、標準パターンのどのフレームに対応するかを求める。図6はフレーム整合の様子を表した図である。ここで、特徴ベクトルが単一ガウス分布に従うと仮定し、標準パターンのそれぞれのフレームについて、対応する学習データの特徴ベクトルの平均値ベクトルおよび共分散を求める。これが標準パターンとなる。本実施の形態1では単一ガウス分布とするが混合ガウス分布としても良い。

#### 【0037】

以上を、N種の雑音すべてについて行う。最終的には、N種の雑音を重畠したデータにたいする平均値ベクトル及び共分散行列が求まる。

#### 【0038】

このようにして、多数の雑音サンプルをクラスタリングしてそれぞれのクラスから1つずつ学習に用いる雑音を選択することにより決定される複数の雑音を重畠した学習用音声サンプルを用いて音響モデルの学習を行うことにより、少ない雑音サンプルで学習でき、しかも偏り無くさまざまな雑音を広くカバーすることができるため、未知雑音環境下においても高い認識性能が得られる音響モデルを作成できる。

#### 【0039】

##### (実施の形態2)

実施の形態2は、実施の形態1で作成された音響モデルを用いて、入力された音声を認識する音声認識装置について説明する。図8に、本発明の実施の形態2

における音声認識装置の動作フローチャートを示し、以下に説明する。

【0040】

図8において、音響モデル108は実施の形態1で作成されたものを用いる。

【0041】

まず最初に、未知入力音声をフレームごとに音響分析し、特徴パラメータを抽出する(S61)。特徴パラメータは音響モデル作成時と同じものである。

【0042】

認識対象語彙辞書109には、認識対象となる語彙の音節列が表記されている。この認識対象語彙辞書にしたがって、あらかじめ作成されている音節単位の音響モデル108を接続し、これと、上記未知入力音声の特徴パラメータ時系列とをDP照合する(S62)。その結果、最も類似度の大きかった語彙を認識結果として出力する。

【0043】

このようにして動作する音声認識装置では、音響モデルが、多数の雑音サンプルをクラスタリングすることにより決定される複数の雑音を重畠した学習用音声サンプルを用いて学習されているため、未知雑音環境下においても高い認識性能が得られる。

【0044】

以下に、認識実験について説明する。

【0045】

本発明の効果を検証するため、本実施の形態によって得られる音響モデルを用いて音声認識実験を行った。評価データは10名分の100地名音声データとした。評価データに、学習用に用いなかった雑音サンプルを重畠し、100単語認識を行った。学習用複数雑音サンプルは、「川」、「マークII」、「データシヨウ会場」、「事業部」、「幼稚園」である。

【0046】

評価データへ重畠する雑音サンプルは、クラス1からは「音楽」、クラス2からは「マジエスタ」、クラス3からは「地下鉄」、クラス4からは「事務所」、クラス5からは「東京駅」とした。また、まったく未知の雑音として、「道路」

(道路わきで収録した騒音)、「テレビCM」(テレビのコマーシャル音を収録したもの)をそれぞれ評価データに重畠して、単語認識実験を行った。

## 【0047】

また、本発明の対照実験として、従来例に相当する、1種類の雑音のみ(「マークII」)で学習した音響モデルを用いた単語認識実験も同様に行った。

## 【0048】

その結果を(表1)に示す。

## 【0049】

## 【表1】

100 単語認識率(%)

評価データノイズ		100 単語認識率(%)							
		音楽	マジェスタ	地下鉄	事務所	東京駅	道路	テレビCM	
学習データノイズ	マークII	(1)	48.2	94.8	88.8	76.7	77.7	92.0	58.2
クラス1~5	川、マークII、データショウ会場、事業部、幼稚園	(2)	77.1	92.9	92.7	90.5	91.3	94.0	74.1

## 【0050】

(表1)から、(1)において、学習時の雑音と認識時の雑音が同じクラス内(クラス2同士)の場合、高い認識性能が得られる(マジェスタ94.8%)が、他のクラスに属する雑音環境下では認識性能が悪い。これに対し、本発明の実験(2)では、クラス2以外のすべてのクラスで(1)よりも高い認識性能を示している。さらに、未知雑音環境に対する実験では、「道路」に対しても「テレビCM」に対しても、本発明の方が認識性能が高いことがわかる。

## 【0051】

このことから、本発明は未知雑音に関しても高い認識性能が得られるといえる。

## 【0052】

なお、本実施の形態では、クラスタリングによって選択したN個の学習用雑音サンプルをそれぞれ学習用音声サンプルに重畠して、音響モデルの各状態を单一ガウス分布として学習を行ったが、雑音ごとに別々にN個のガウス分布の混合分

布としても良い。また、単一ガウス分布で表されるN個の音響モデルを学習して、認識時にN個の音響モデルと照合を行い、最も類似度の高い音響モデルに対するスコアを最終スコアとしても良い。

【0053】

【発明の効果】

以上のように本発明は、クラスタリングにより選択した雑音を学習データに重畳することにより、比較的少ない複数の雑音で学習できるため効率的であり、しかも偏り無く、広く雑音環境をカバーできるため、未知の雑音に対しても高い認識性能が得られる音響モデルが作成できる。

【図面の簡単な説明】

【図1】

本発明の実施の形態1における音響モデル作成装置の構成図

【図2】

本発明の実施の形態1における学習用複数雑音サンプルの選択方法を表すフローチャート

【図3】

本発明の実施の形態1における音響モデルの作成方法を表すフローチャート

【図4】

多数雑音サンプルの詳細を説明する図

【図5】

樹形図

【図6】

フレーム整合のイメージ図

【図7】

本発明の実施の形態2における音声認識方法を表すフローチャート

【図8】

従来例の音響モデル作成装置の構成図

【図9】

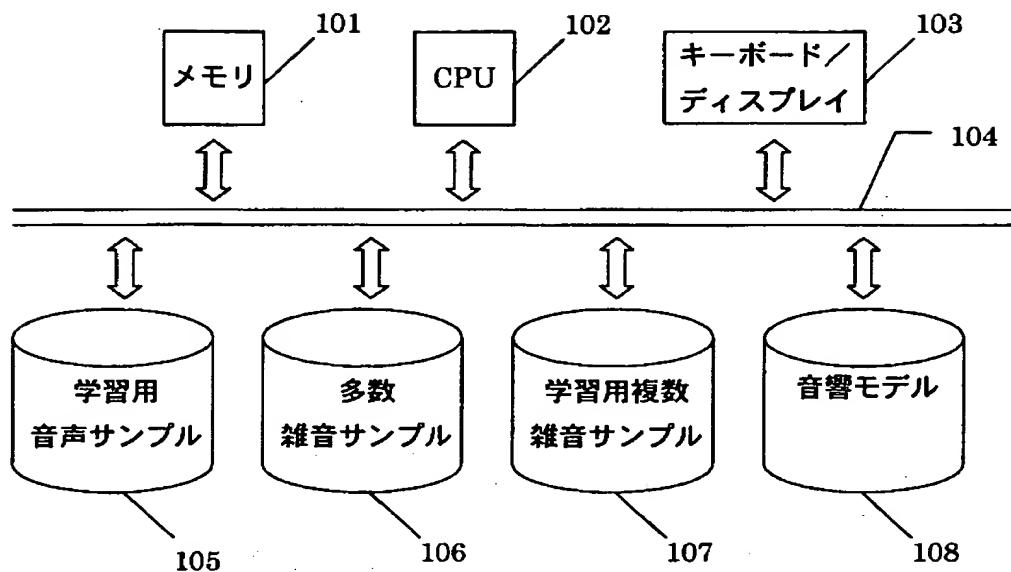
従来例の音響モデルの作成方法を表すフローチャート

【符号の説明】

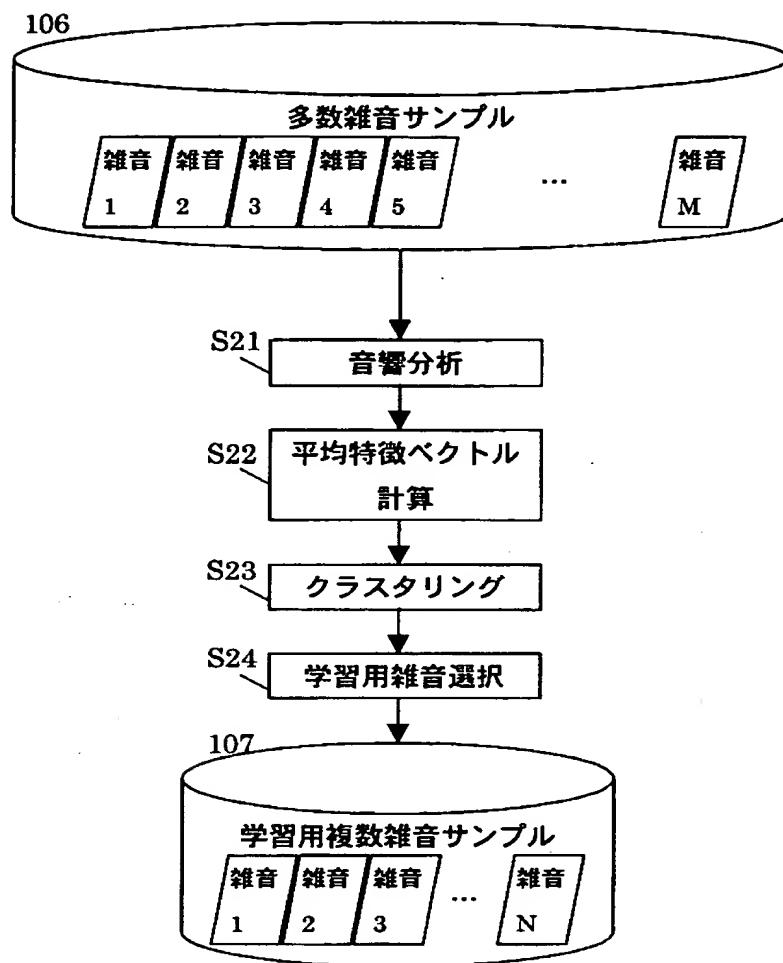
- 101 メモリ
- 102 CPU
- 103 キーボード／ディスプレイ
- 104 CPUバス
- 105 学習用音声サンプル
- 106 多数雑音サンプル
- 107 学習用複数雑音サンプル
- 108 音響モデル
- 109 認識対象語彙辞書

【書類名】 図面

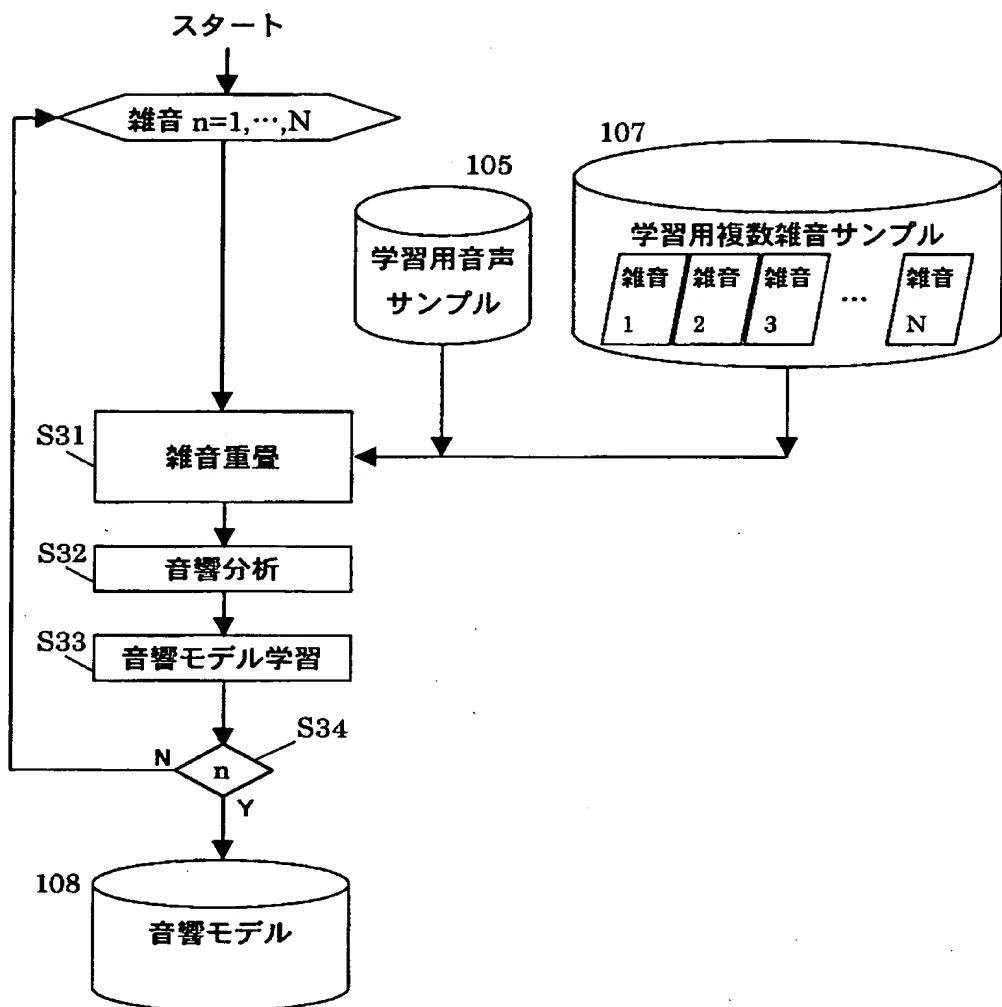
【図1】



【図2】



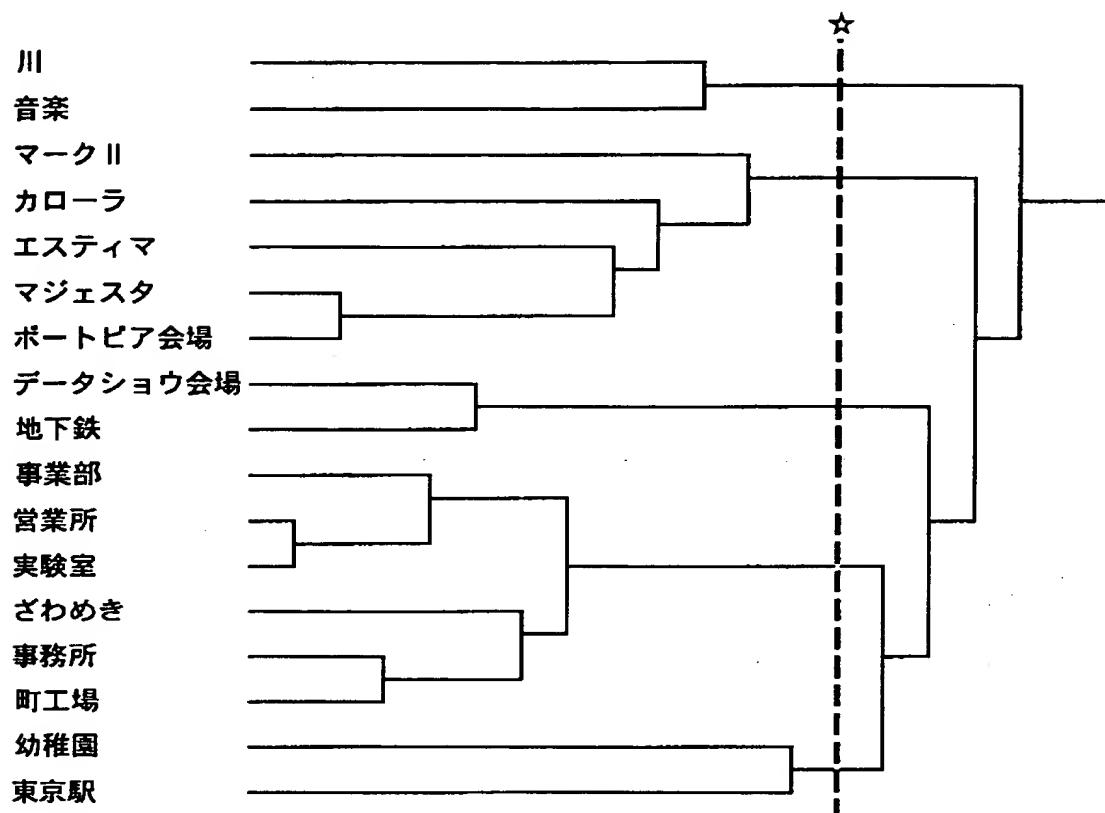
【図3】



【図4】

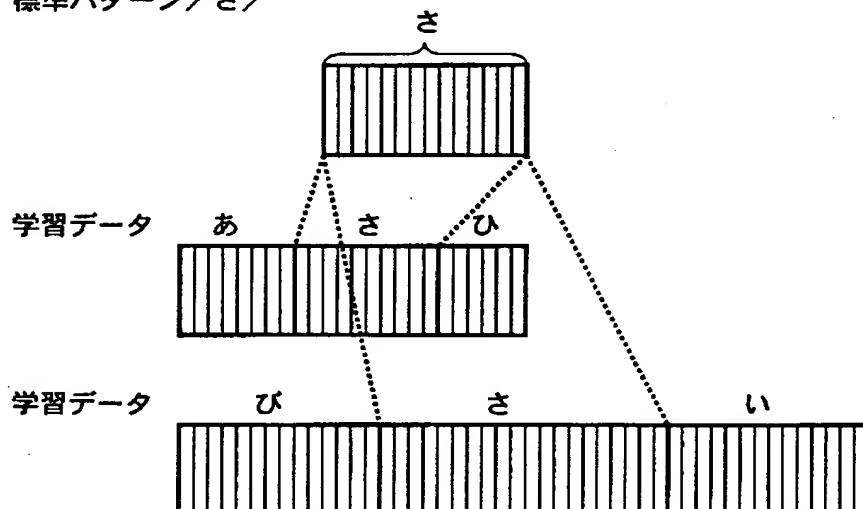
ノイズ名	備考
川	川のせせらぎ音
音楽	ポップス
マークII	車(マークII)走行時雑音
カローラ	車(カローラ)走行時雑音
エスティマ	車(エスティマ)走行時雑音
マジェスタ	車(マジェスタ)走行時雑音
ポートピア会場	ポートピア会場騒音
データショウ会場	データショウ会場騒音
地下鉄	地下鉄騒音
事業部	事業部騒音
営業所	営業所騒音
実験室	技研実験室
ざわめき	効果音CD
事務所	事務所騒音
町工場	輪転機
幼稚園	幼稚園騒音
東京駅	東京駅八重洲中央口騒音

【図5】

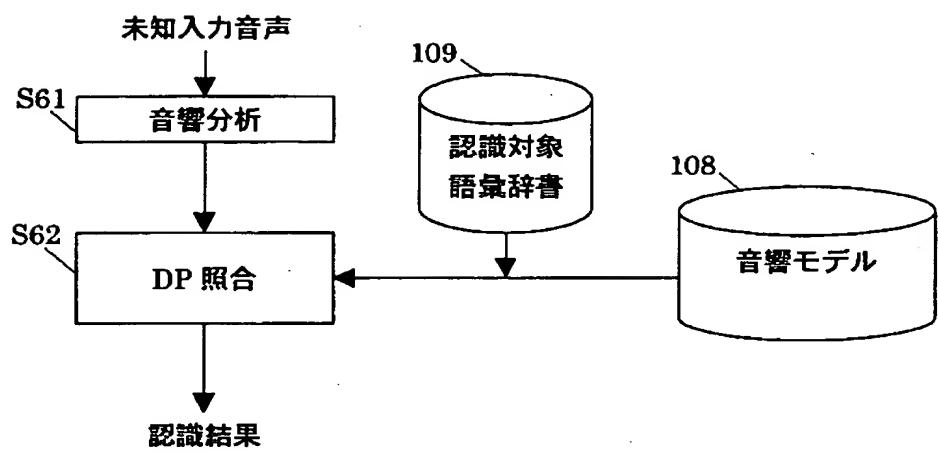


【図6】

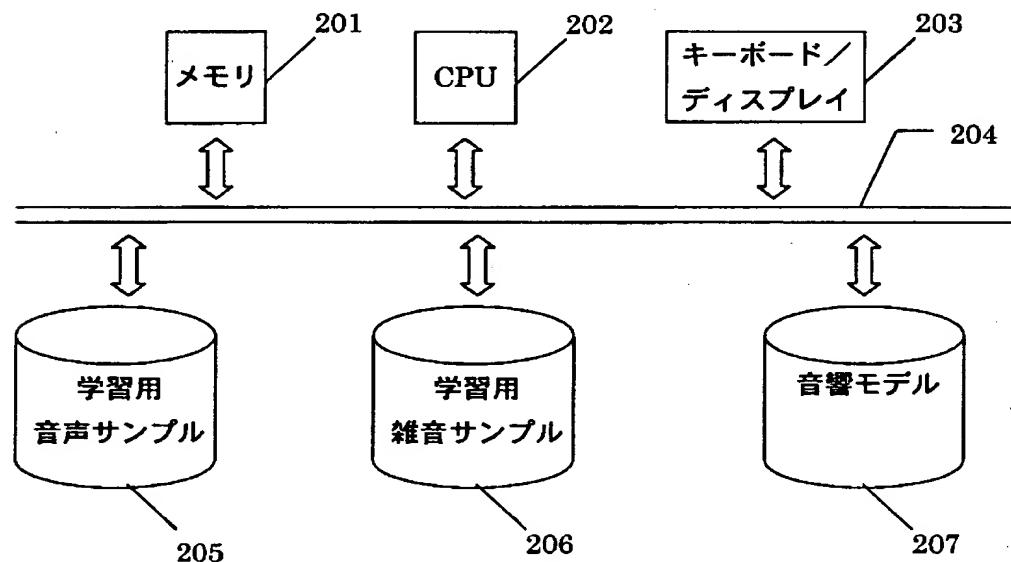
標準パターン／さ／



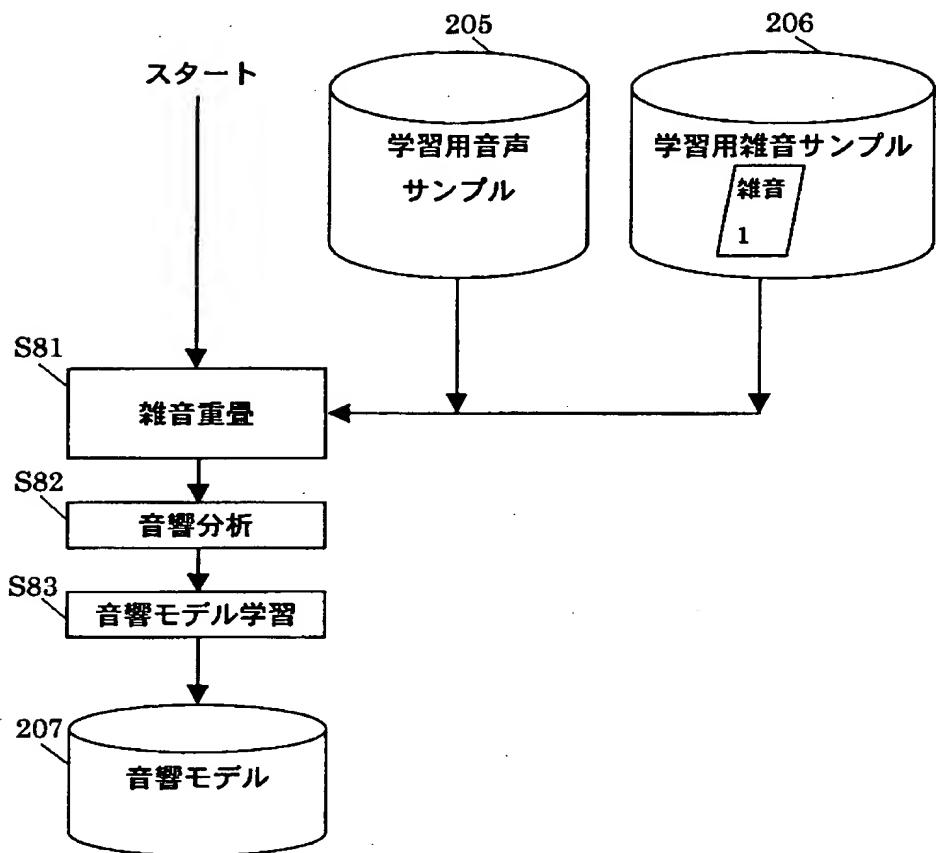
【図7】



【図8】



【図9】



【書類名】 要約書

【要約】

【課題】 本発明は音声認識用音響モデル作成技術に関するものであり、雑音環境下で高い認識性能が得られる音響モデルを作成することを目的とする。

【解決手段】 あらかじめ多数の雑音サンプルをクラスタリングし、各クラスから1つずつ雑音サンプルを選択して学習用複数雑音サンプルとし、学習用音声サンプルにその学習用複数雑音サンプルをそれぞれ重畠したデータを用いて音響モデルを作成するものである。これにより、クラスタリングにより選択した雑音を学習データに重畠することにより、少ない雑音サンプルで効率的に学習でき、しかも偏り無く広く雑音環境をカバーできるため、未知の雑音に対しても高い認識性能が得られる音響モデルが作成できる。

【選択図】 図3

出願人履歴情報

識別番号 [000005821]

1. 変更年月日 1990年 8月28日

[変更理由] 新規登録

住 所 大阪府門真市大字門真1006番地

氏 名 松下電器産業株式会社